

scikit-learn ゼミ

1.3.5. Tips on Practical Use

小野寺喜行

Stochastic Gradient Descent

SGDはデータのスケールに影響を受けやすい



データに応じてスケールを調整することが推奨される

調整方法の例

- ・入力ベクトル X を $[0,1]$ か $[-1,+1]$ にする
- ・平均0かつ分散1に標準化

※テストベクトルも同じスケールにする必要がある

※固有のスケール(単語頻度、指標機能など)を持つ場合は必要ない

StandardScaler

StandardScalerによって簡単にスケールの調整が可能

```
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
scaler.fit(X_train) # Don't cheat - fit only on training data
X_train = scaler.transform(X_train)
X_test = scaler.transform(X_test) # apply same transformation to test data
```

- fit(): 後のスケーリングに使われる平均と標準偏差を計算
- transform(): センタリングとスケーリングによって標準化を実行

GridSearchCV

- ・適切な正規化項の導出はGridSearchCVが最も良い

※範囲: `10.0**-np.arange(1,7)`

Stochastic Gradient Descent

- ・約 10^6 個のトレーニング・サンプルでSGDが収束する
- ・反復する数のための最適な推測: $n_iter = \text{np.ceil}(10^{**}6/n)$
(nはトレーニングのサイズ)

SGD to features extracted using PCA

- ・PCAで引き抜かれた特徴にSGDを適用するとき
 - トレーニング・データのL2ノルムの平均が1になる様に、いくつかの定数 c によって特徴の値を調整する